



October 16, 2024

Sent via ECF

Hon. Ona T. Wang
Daniel Patrick Moynihan
United States Courthouse
500 Pearl Street
New York, NY 10007-1312

Orrick, Herrington & Sutcliffe LLP
The Orrick Building
405 Howard Street
San Francisco, CA 94105-2669

+1 415 773 5700
orrick.com

Annette Hurst

E ahurst@orrick.com
D +1 415 773 4585
F +1 415 773 5759

RE: *The New York Times Co. v. Microsoft Corp., et al.* No. 1:23-cv-11195
New York Daily News, LP, et al. v. Microsoft Corp., et al., No. 1:24-cv-3285

Dear Magistrate Judge Wang:

On behalf of all parties in the above-referenced actions and pursuant to the Court's September 13, 2024 Order re Motions to Compel & Status Conference ("Order"), ECF No. 202, counsel for Microsoft hereby submits the proposed agenda for the October 30, 2024 Status Conference.

The parties met and conferred in an effort to streamline the issues before the Court. All items identified by any party for resolution have been included in the list below and a brief table of contents has been included for ease of review.

Table of Contents of Arguments

| | |
|---|----------|
| I. Updates to Prior Disputes Addressed at the September 12, 2024 Conference..... | 3 |
| A. Dkt. 128 – The Times's motion to compel OpenAI to produce documents concerning large language models developed before 2022, specifically GPT (released in 2018), GPT-2 (released in 2019), and GPT-3 (released in 2020)) | 3 |
| B. Dkt. 141 – The Times's motion to compel OpenAI to produce documents in response to RFP 6 (communications regarding licensing agreements) and RFP 11 (documents concerning OpenAI's formation of a for-profit entity). | 4 |
| C. Dkt. 204 –The Times's Motion to Compel OpenAI to Produce Documents from Additional Custodians | 7 |

| | | |
|------------|--|-----------|
| D. | Dkt. 221 – The Times’s motion to compel Microsoft to re-review the documents it designated as “Highly Confidential – Attorneys’ Eyes Only” | 8 |
| E. | Dkt. 228 – The Times’s motion to compel OpenAI to apply additional search terms to its custodians’ ESI..... | 9 |
| F. | Model Inspection, Training Data Inspection, Log Data, and Source Code..... | 11 |
| II. | New Disputes that Have Materialized Since the September 12, 2024 Conference..... | 13 |
| A. | Issues Raised by Defendants | 13 |
| 1. | The Times’ Responses to Microsoft RFPs and Interrogatories..... | 13 |
| 2. | The Times’ ESI Custodians..... | 14 |
| 3. | Date Ranges of The Times’ Searches for Responsive Documents | 15 |
| 4. | The Times’ Responses to OpenAI’s RFPs and Interrogatories..... | 16 |
| B. | Issues Raised by Plaintiffs..... | 23 |
| 1. | Microsoft’s ESI Custodians | 23 |
| 2. | Microsoft Responses to The Times’s RFPs | 24 |
| 3. | The Scope of OpenAI’s Products Subject to Discovery..... | 26 |
| 4. | OpenAI’s Responses to RFPs and Interrogatories | 28 |

I. UPDATES TO PRIOR DISPUTES ADDRESSED AT THE SEPTEMBER 12, 2024 CONFERENCE

A. Dkt. 128 – The Times’s motion to compel OpenAI to produce documents concerning large language models developed before 2022, specifically GPT (released in 2018), GPT-2 (released in 2019), and GPT-3 (released in 2020))

News Plaintiffs’ Statement: The parties have made progress on this dispute, which includes OpenAI agreeing to produce additional documents and training data relating to GPT, GPT-2, and GPT-3.¹

One dispute remains. Because OpenAI has now agreed these models should be subject to discovery, OpenAI should also respond to Interrogatories with respect to these models. The Times’s motion at Dkt. 128 focused on The Times’s RFPs, but only because OpenAI had not yet responded to any Interrogatories. When OpenAI later responded to The Times’s First Interrogatories, the same dispute regarding these earlier models re-surfaced, and the parties agreed that any resolution of the motion filed at Dkt. 128 would resolve this same dispute with respect to the Interrogatories. See August 14, 2024 Email from OpenAI’s counsel (“OpenAI agrees that the motion pending before the Court will likely resolve the dispute as to the Text Generation AI Models at issue.”). Because the parties have now agreed that GPT, GPT-2, and GPT-3 will be subject to discovery, the News Plaintiffs request that OpenAI be ordered to supplement its Interrogatory responses to respond as to these models, to the extent any response would differ as applied to these models. The News Plaintiffs can work with OpenAI to determine which Interrogatory responses require supplementation.

OpenAI’s Statement: OpenAI has resolved most of this dispute and remains willing to confer on the remainder.

First, as to the scope of document production, OpenAI has agreed to reasonable search terms that are likely to capture custodial documents relating to GPT, GPT-2 and GPT-3. OpenAI is not refusing to produce those documents on the basis that they relate to those models, despite its objections regarding relevance (since these models were not used for ChatGPT). OpenAI has also agreed to produce or otherwise make available the following: (1) training data for GPT, GPT-2 and, GPT-3; and (2) non-custodial documents related to GPT-2 and GPT-3, as it has for other models, to the extent they are available and can be located after a reasonable search. This agreement encompasses the documents News Plaintiffs seek. Based on the above, OpenAI does not understand the parties to have a dispute as to scope.

¹ The parties still have disputes about the exact set of search terms to use to capture documents relating to these models, as reflected below in the section about Dkt. 228.

Second, as to The Times’ interrogatories, prior to October 15, 2024, The Times had not raised this issue with OpenAI, the parties had not discussed the scope, and OpenAI had not had the opportunity to consider the request. OpenAI is willing to assess whether any specific interrogatories need to be supplemented, but it does not believe they all do. For example, Interrogatory No. 1 asks OpenAI to, among other things, “Identify all individuals with knowledge of (i) Defendants’ Training Datasets, including the existence of Times Content and other copyrighted content therein[.]” This information is not relevant for GPT if, e.g., The Times has no reason to believe its information was included in training data used for GPT. See A. Radford et al., “Improving Language Understanding by Generative Pre-Training,” at 4 (describing GPT as having been trained on “unpublished books from a variety of genres including Adventure, Fantasy, and Romance”). The News Plaintiffs’ request that the Court order OpenAI to supplement its interrogatory responses is thus premature. OpenAI is willing to meet and confer on this issue to determine whether or not a dispute exists.

B. Dkt. 141 – The Times’s motion to compel OpenAI to produce documents in response to RFP 6 (communications regarding licensing agreements) and RFP 11 (documents concerning OpenAI’s formation of a for-profit entity).

i. RFP 6: (communications regarding licensing agreements)

News Plaintiffs’ Statement: OpenAI has agreed to produce some communications between OpenAI and third parties with whom OpenAI had a term sheet or an executed data access agreement. But two disputes remain regarding RFP 6.

First, OpenAI should also be ordered to produce communications regarding *potential* licensing deals that never materialized into a signed term sheet or agreement. Such communications are relevant for the same reasons as communications regarding successful negotiations. As one example, OpenAI’s willingness to pay a third party a certain amount for content is relevant regardless of whether those negotiations culminated in a signed agreement. Contrary to what OpenAI claims below, parties can have substantive discussions about a licensing agreement without reaching a signed term sheet. Indeed, OpenAI seeks to exclude communications regarding negotiations with The Times.

Second, OpenAI should also be ordered to produce *internal* communications relating to actual and potential licensing agreements, not just communications with third parties. OpenAI has argued that internal communications are likely to be “entirely privileged”, but that assertion is dubious, including because The Times has collected and is producing internal communications relating to its own potential licensing deals with Microsoft and OpenAI. Moreover, OpenAI’s internal communications may be even more informative than its communications with third

parties, since internal communications will reveal OpenAI's actual views on the licensing deals, as opposed to the view that OpenAI chose to externally project. OpenAI below includes a compromise offer made for the first time in this statement, which the News Plaintiffs will evaluate.

OpenAI's Statement: OpenAI has already produced most of the relevant data access agreements.² For the remainder, OpenAI is diligently addressing third party confidentiality concerns.

OpenAI has also proposed a reasonable scope of production for communications related to those data access agreements.

First, OpenAI has agreed to produce non-privileged communications between OpenAI and third parties with whom OpenAI had an executed data access agreement or term sheet, where the third party does not object to the production of their information. To the extent these communications are not privileged and discuss financial terms, releases or covenants not to sue, term and termination provisions, and scope of use, OpenAI will produce them.

Second, OpenAI has agreed to produce non-privileged communications with third parties related to negotiations where OpenAI and the third party have a term sheet. This is a reasonable scope because it identifies the set of documents where OpenAI had partnership discussions that had some possibly relevant substantive discussion. The News Plaintiffs have neither suggested that this scope is inadequate, nor explained why any additional search would be required or proportional. Nor have the News Plaintiffs offered an alternative. They have not explained when a discussion with a third party should be considered to be a communication about a "potential" agreement. OpenAI has made a reasonable proposal as to how to identify such communications by determining whether there is a term sheet that discusses a relevant potential agreement.

Third, OpenAI has already designated several custodians that are knowledgeable about OpenAI's data access agreement negotiations. And it has proposed performing a reasonable search for non-privileged documents reflecting or discussing access agreement negotiations (whether or not those negotiations resulted in an executed agreement) in the files of the currently designated custodians. To be clear, that reasonable search would include internal discussions. However, because OpenAI lawyers were heavily involved in negotiations, OpenAI anticipates a significant privilege burden were it to agree to review all internal communications. As a compromise, OpenAI has proposed that its search will exclude internal legal communications. In other words, OpenAI would not review, log, or produce any internal communication about negotiations that was received or sent by an OpenAI lawyer. This approach addresses the significant privilege review burden that Plaintiffs' approach would otherwise impose on OpenAI.

² OpenAI does not concede that any data agreements produced are licenses.

OpenAI's proposed scope of search and production is reasonable. However, even if Plaintiffs disagree, the dispute appears premature given that Plaintiffs have not yet considered the documents OpenAI is producing.

ii. **RFP 11: (documents concerning OpenAI's formation of a for-profit entity)**

News Plaintiffs' Statement: OpenAI has agreed to produce documents about OpenAI's decision to transition to a for-profit entity that also discuss or address OpenAI GPT large language models or GPT products. With respect to Microsoft, OpenAI has agreed to produce documents about OpenAI's decision to transition to a for-profit entity that relate to Microsoft (including communications with Microsoft) to the extent such documents also discuss or address OpenAI GPT large language models or GPT products.

Certain disputes remain, as outlined below.

The News Plaintiffs object to OpenAI's proposal to limit these productions to documents that "also discuss or address OpenAI GPT large language models or GPT products." OpenAI has clarified in meet and confers that, if a document "says nothing about OpenAI's products or services at all, then OpenAI will not produce the document." The problem with this limitation is that a document could address relevant issues without explicitly mentioning a specific model or product—for example, an email could discuss OpenAI's plans to secure investments without explicitly mentioning "GPT", and OpenAI would admittedly withhold such a document. OpenAI's position also makes no sense given that, at this time of its for-profit transition (2019), OpenAI's primary line of business was its creation of GPT models and products. As a result, any document regarding the for-profit transition would be relevant to those models and products, regardless of whether the document explicitly mentioned the models and products.

Without limiting productions to documents that mention specific models and products, OpenAI should be ordered to produce documents relating to how the transition impacted OpenAI's business strategy more broadly—including, for example, how OpenAI would secure investments, generate revenue, and otherwise engage with the marketplace, all of which are relevant to, at a minimum, the fair use analysis, including the purpose and character of the use.

OpenAI below includes a revised offer made for the first time in this statement, which the News Plaintiffs will evaluate.

OpenAI's Statement: There should be no dispute here. OpenAI is willing to produce documents regarding its decision to create a for-profit entity. It simply seeks to limit the scope of the request to documents relevant to the case.

OpenAI has agreed to conduct a reasonable search of its designated custodians and produce documents relating to OpenAI's creation of a for-profit entity (regardless of whether those documents are communications with Microsoft) to the extent those documents address the operation of OpenAI's relevant business or, as a general matter, its ability to raise capital. Given that documents addressing a change of corporate structure are likely to overwhelmingly involve lawyers, OpenAI has proposed that, as part of the search, OpenAI would not review, log, or produce any internal communication that was received or sent by an OpenAI lawyer.

OpenAI's proposal is reasonable and reflects its concern that Plaintiffs' request, as worded, would sweep in numerous irrelevant documents—a fact Plaintiffs have acknowledged in the parties' conferrals. For example, corporate structures are designed for a variety of reasons, such as to minimize tax or compliance burdens. Plaintiffs acknowledge that none of those reasons are relevant here, yet have refused to frame this request to exclude such documents and target the ones actually relevant to their claims. OpenAI's proposal does just that. Moreover, OpenAI's custodians include the relevant high-level executives and decision makers (e.g., its CEO and President) on this topic, who are likely to have responsive documents to the extent they exist. If there are additional substantive topics that Plaintiffs believe should be covered by this request, OpenAI is open to conferring on those issues.

C. Dkt. 204 –The Times's Motion to Compel OpenAI to Produce Documents from Additional Custodians

News Plaintiffs' Statement: At the September 12 hearing, the Court ordered that OpenAI preserve documents for the eight custodians at issue in The Times's August 12 motion to compel, and the Court instructed the parties to revisit this dispute after OpenAI completed additional custodial document productions. *See* Sept. 12, 2024 Hearing Tr. at 101:15-25 (“Hold onto it. See what you get [in] the production that you're getting from the custodians already agreed to and then see if there is a need to further search those additional custodians. I think this is a matter that we don't have to decide once and for all today so long as the documents are preserved.”). The News Plaintiffs planned to review forthcoming productions and revisit this issue with OpenAI in advance of the October 30 conference.

But after the September 12 hearing, OpenAI waited until October 14 to make any additional document productions that could potentially be useful for this dispute. For this reason, The Times respectfully reiterates its request for OpenAI to produce documents from the eight custodians: James Betker, William Fedus, Daniel Kokotajlo, Brad Lightcap, Ryan Lowe, William Saunders, Ilya Sutskever, and Wojciech Zaremba. The Times's August 12 motion cites and attaches documents produced by OpenAI produced (earlier in discovery) to support The Times's request. Dkt. 204 (motion to compel). OpenAI below promises that document productions are forthcoming, and the News Plaintiffs will review them. But given the approaching

December 20 close of fact discovery, the News Plaintiffs respectfully submit that now is an appropriate time to resolve this dispute, so there is sufficient time for OpenAI to produce these custodians' documents.

OpenAI's Statement: Plaintiffs acknowledge the Court's instruction that the parties should revisit this dispute after Plaintiffs have reviewed the productions from OpenAI's designated custodians. Since the September hearing, OpenAI has been working diligently to review tens of thousands of documents from its 24 agreed-upon custodians, including the 10 additional custodians it agreed to add last month. Contrary to Plaintiffs' suggestion above, OpenAI has made multiple productions of documents since the September 12 status conference and is preparing to make additional productions containing thousands of custodial documents before October 30. Despite knowing that those productions are imminent, Plaintiffs have renewed their demand for additional custodians without reason to think the individuals they identify possess relevant, responsive, and non-cumulative information.

Plaintiffs' insistence on parroting the same demand for custodians at each successive discovery conference—before they have reviewed the documents that OpenAI has agreed to produce—is burdensome and inefficient. Plaintiffs should complete their review from OpenAI's agreed-upon custodians and meet and confer with OpenAI about any gaps before again burdening the Court with this issue.

D. Dkt. 221 – The Times's motion to compel Microsoft to re-review the documents it designated as “Highly Confidential – Attorneys' Eyes Only”

Joint Statement: Since the September 12 hearing, the parties have continued to meet and confer regarding this issue. Microsoft re-designated approximately 9,500 previously AEO documents as Confidential. 65% of Microsoft's documents are now designated AEO.

News Plaintiffs' Statement: Although the News Plaintiffs appreciate the progress that has been made on this issue, many documents remain improperly designated. The Times has provided Microsoft with several examples, yet Microsoft contends that documents containing, for example, news updates and publicly available articles are properly designated as AEO because they were collected from certain “high level custodians” and “C-suite executives.” News Plaintiffs disagree that documents are necessarily AEO because they were collected from high level custodians. News Plaintiffs remain open to amendments to the Protective Order that would alleviate Microsoft's concerns regarding the confidentiality designation of documents produced to OpenAI, which it claims is a competitor. The parties met and conferred regarding this issue on October 16 and reached a tentative verbal agreement to amend the Protective Order to permit certain specified Times executives to view Microsoft documents marked AEO, provided those documents do not relate to licensing and negotiations over license agreements. News Plaintiffs

await final confirmation from Microsoft that Microsoft is amenable to this resolution.

Microsoft's Statement: While the parties have conferred in good faith regarding a proposed amendment to the Protective Order, the News Plaintiffs have indicated that they would prefer to require Microsoft to further re-review its documents rather than accepting a modification to the Protective Order. Accordingly, Microsoft undertook extensive efforts to re-review and re-designate categories of documents produced to date and to provide its document reviewers with detailed guidance for reviewing documents for confidentiality going forward. As a result, the rate of designation of AEO documents in Microsoft's production dropped to 73% upon re-review of documents produced to date, and dropped to 32% in Microsoft's most recent production, which has resulted in an overall AEO percentage of 65%. This rate is expected and appropriate given that the documents requested by the News Plaintiffs involve commercially and competitively sensitive information regarding emerging technologies, requested from multiple high-level executives, in a case involving both Microsoft's competitor and allegations of competitive harm by the News Plaintiffs. The parties are continuing to meet and confer regarding this issue, including with respect to potential amendments to the Protective Order and any further re-review and re-designation is unnecessary and unwarranted.

E. Dkt. 228 – The Times's motion to compel OpenAI to apply additional search terms to its custodians' ESI

News Plaintiffs' Statement: Since the September 12 hearing, OpenAI and the News Plaintiffs have been exchanging search term proposals, and have made some progress.

But a wide gap remains. A gating issue is that the parties have very different views on how many total documents OpenAI should collect and review. The News Plaintiffs' most recent proposal generated 632,537 additional document hits, while OpenAI's most recent proposal generated 195,362 additional document hits. The parties are continuing to exchange proposals and to refine proposed terms, but full agreement on search terms appears unlikely given this gating dispute.

The News Plaintiffs request that the Court order OpenAI to run search terms that hit on approximately 550,000 to 600,000 additional documents. Once the Court provides guidance on this gating dispute, the parties should be able to negotiate the specific terms to be used and to work through any issues regarding specific terms. The News Plaintiffs' proposal is appropriate number given the breadth of issues in this case, as outlined in The Times's August 27 motion to compel additional search terms. *See* Dkt. 228. OpenAI has so far produced only 35,000 documents, which this Court described as "not that many anymore." *See* Sept. 12 Hearing Tr. at 31:11 (describing the 42,000 documents that OpenAI produced in the *Authors Guild* class action). A substantially larger document production is warranted and proportional,

particularly because this case implicates over ten million allegedly misappropriated works. The recent consolidation of The Times's case with the Daily News case provides additional support for the News Plaintiffs' proposal, since this proposal now applies to two cases, not just one.

The News Plaintiffs also request that any target amount set by the Court also and separately apply to Microsoft's custodial ESI searches. The News Plaintiffs and Microsoft have not made much progress because Microsoft only last week made a counter-proposal for search terms. After The Times made three separate proposals in which Microsoft rejected each and every proposed term (on July 2, July 29, and August 26) The Times on September 13 (one day after the hearing) asked Microsoft to finally make a counter-proposal for search terms, consistent with OpenAI's agreement to do so at the hearing. But Microsoft did not make any counter-proposal with hit counts until October 10, roughly one month after the hearing.

OpenAI's Statement: OpenAI has reviewed almost 400,000 documents to date and is currently in the process of reviewing tens of thousands of additional documents. OpenAI has produced tens of thousands of documents (over 629,000 pages) and over 56 gigabytes of data to the News Plaintiffs, with more productions expected soon. OpenAI has also provided over 175 terabytes of training data to the News Plaintiffs for inspection.

OpenAI's document review and proposals have been more than reasonable. As OpenAI already explained, there are significant costs associated with reviewing documents. See Dkt. No. 233-2 (declaration from OpenAI's discovery vendor discussing the costs of review). While News Plaintiffs have suggested this is appropriate given "over ten million allegedly misappropriated works," News Plaintiffs have not explained how the number of allegedly misappropriated works implies OpenAI should review more emails or Slack messages. Specifically, the News Plaintiffs' proposal does not account for what documents have been produced or their substantive nature, nor what documents OpenAI has already agreed to produce. Instead, the News Plaintiffs appear to be setting a target solely for the sake of setting a target. Given that the Plaintiffs have not identified any substantive gaps in OpenAI's productions to date or suggested OpenAI's terms are deficient, their proposal imposes a burden without a basis in Rule 26.

OpenAI also has significant concerns regarding the scope of the terms the News Plaintiffs have suggested. For example, several of their terms hit on tens of thousands of documents. The News Plaintiffs, for example, have proposed the term "The Times" without any additional limitation. This hits on over 34,000 documents when family members are included. This is not a reasonable term given that it would capture, for example, documents that include the phrase "the times it ran" or "I have the times on my calendar." Another term, relating to "clickthrough," hits on over 64,000 documents with family members. Even assuming every document is responsive, the News Plaintiffs have not explained why they need 64,000 documents

on clickthrough rates. OpenAI has provided reasonable search terms, and the News Plaintiffs have not identified any deficiencies with those terms.

OpenAI is actively reviewing documents and more productions are forthcoming. Although the News Plaintiffs attempt to minimize OpenAI's production of around 35,000 documents to date, this number, if anything, is a consequence of Plaintiffs' overbroad terms OpenAI nonetheless compromised by agreeing to run and review. As of October 15, 2024, the News Plaintiffs have asked OpenAI to run terms leading to the review of 632,000 documents and OpenAI has countered with narrowed terms and a reasonable proposal to review 195,000 documents. Plaintiffs' current proposed terms are overbroad and likely to result in OpenAI reviewing hundreds of thousands of nonresponsive documents. OpenAI objects to those terms as unreasonable, burdensome, and not proportional to the needs of the case.

Even for documents which have been reviewed and preliminarily determined to be responsive, OpenAI engages in additional reviews (e.g., for privilege, and PII) prior to making productions. That takes time; but OpenAI is working through the review and production process steadily and in good faith. To the extent there remains any dispute, the dispute should be raised only after OpenAI has produced relevant, non-privileged documents responsive to its proposed search terms and the News Plaintiffs identify an alleged gap. Instead of reasonably addressing discovery logistics, Plaintiffs' proposal instead threatens to impose on OpenAI even more burdensome reviews of overbroad terms likely to lead to even more non-responsive documents.

Microsoft's Statement: While the parties have met and conferred multiple times regarding search terms over the past several months, each iteration of the Times' proposed search terms returned *more than 3 million parent-level items*. After disclosing hit reports on each proposal from the Times, Microsoft provided the Times with detailed feedback and suggestions on methods for narrowing the search terms to a set that is reasonable in scope and appropriately tailored to the case. To date, the Times has not implemented these suggestions and has not proposed a set of search terms that are drawn to return a reasonable and not unduly burdensome set of documents. Nevertheless, Microsoft provided a counterproposal and hit report that has substantially narrowed the number of parent-level items. The parties are continuing to meet and confer regarding these search terms.

F. Model Inspection, Training Data Inspection, Log Data, and Source Code

News Plaintiffs' Statement: At the September 12 hearing, the Court ordered the parties to report by September 27, 2024, on progress on the inspection protocols under which the News Plaintiffs' will inspect Defendants' training data and models. *See* Sept. 12, 224 Hearing Tr. at 120:8-14.

- News Plaintiffs and OpenAI have reached agreement on a Training Data Inspection Protocol (Dkt. 254), and News Plaintiffs will begin their initial inspection of OpenAI's training data on October 16, 2024.
- As set forth in the Parties' Joint Letter to the Court on September 27, 2024 (Dkt. 250), the Parties are at an impasse on a Model Inspection Protocol, and request that the Court address the issue at the October 30 status conference.
- OpenAI has yet to produce any source code or logs for its Generative AI Products and Services, and The News Plaintiffs anticipate continued negotiations over the manner and scope of production of these items.
- News Plaintiffs and Microsoft have not yet reached agreement on any of the outstanding requests in Plaintiffs' Requests for Inspection. Since the September 12 hearing, counsel for News Plaintiffs have been attempting to schedule a meet and confer with Microsoft, which has not yet occurred. News Plaintiffs are requesting that Microsoft produce for its Generative AI Products and Services: (1) post-training data, (2) the Microsoft-specific versions of the GPT-models, (3) source code, and (4) logs. News Plaintiffs anticipate continued negotiations with Microsoft over scope and timing of production of these items leading to the October 30 status conference, but are not optimistic that the Parties will reach a resolution given the positions Microsoft has taken.

OpenAI's Statement: OpenAI has made significant progress on matters of technical discovery.

- OpenAI has coordinated with Plaintiffs on a date to inspect training data (data which OpenAI first made available to Plaintiffs four months ago).
- OpenAI has begun collecting source code for inspection, while continuing to confer with Plaintiffs about the precise scope of the code being requested.
- OpenAI is investigating the technical and practical feasibility of Plaintiffs' request to view user session logs in a way that satisfies user privacy concerns and requirements—all the while inviting Plaintiffs to provide guidance regarding precisely what content within user sessions they are interested in viewing.

The sole issue of dispute presented at the October 30 status conference concerns the amount of additional account credits that Plaintiffs are seeking from OpenAI. The parties articulated their positions in a Joint Letter to the Court dated September 27, 2024 (Dkt. 250). OpenAI is prepared to address this issue with the Court at the upcoming hearing.

Microsoft's Statement: The parties have conferred regarding the News Plaintiffs' requests for inspection and will continue to meet and confer, during which Microsoft anticipates that the parties will be able to work collaboratively to resolve issues. For example, Microsoft has agreed to a source code inspection protocol with the News Plaintiffs and has been working to identify and collect relevant source code which will be made available for the News Plaintiffs' inspection pursuant to that protocol. Microsoft's investigation regarding the requested training data, models, and session data is ongoing.

- With respect to post-training data, Microsoft is continuing to investigate whether there is any relevant training data associated with classifier programs used by Copilot (f/k/a Bing Chat).
- With respect to models, Microsoft is not presently aware of any GPT models in its possession for separate inspection that are not also in OpenAI's possession.
- With respect to session data, Microsoft's investigation regarding the feasibility and burdens associated with inspection is also ongoing. This session information amounts to more than 15 petabytes of data, is archived in cold storage and is, at best, extremely difficult to both access and search. Microsoft looks forward to continuing to meet and confer regarding these requests.

II. NEW DISPUTES THAT HAVE MATERIALIZED SINCE THE SEPTEMBER 12, 2024 CONFERENCE

A. Issues Raised by Defendants

1. The Times' Responses to Microsoft RFPs and Interrogatories

Microsoft's Statement: Microsoft requests that The Times be ordered to produce documents and otherwise respond to Microsoft's RFPs 22, 24, 25, 40, and 47 and Interrogatory No. 2 insofar as they pertain to The Times's prompting of Copilot and ChatGPT to obtain outputs on which The Times relies as facts in support of its Commercial Misappropriation (Count VI) claim in paragraphs 127-135 and 192-197 of the First Amended Complaint ("FAC"), and Trademark Dilution (Count VII) claim in paragraphs 136-142 and 198-204 of the FAC. Counts VI and VII rely upon factual allegations that the *outputs* of Microsoft's Generative AI Products and Services constitute commercial misappropriation of The Times's opportunities (Count VI), or cause tarnishment of The Times's trademark (Count VII). These claims rely on prompting conducted by The Times to establish the elements of, respectively, alleged use by Microsoft to compete with The Times (Count VI) and alleged hallucinations that attribute falsehoods to The Times (Count VII).

The FAC partially discloses or purports to summarize the prompts used to generate these outputs. FAC ¶¶127-35, 136-42. Microsoft therefore sought discovery

about these prompts as well as The Times's other prompting efforts leading up to these prompts, in order to defend the allegations of competitive use by Microsoft and to defend the allegation that consumers will perceive outputs in a particular way. Either one of these elements requires an understanding of how The Times prompted the generative AI tools to produce the results upon which its claims rely. Nevertheless, The Times has refused to produce the discovery, claiming privilege. The Times's partial disclosure and affirmative reliance upon the alleged prompts and outputs in Counts VI and VII of the FAC waives any privilege and/or work product protection under Federal Rule of Evidence 502(a). There are no other facts supporting the output-based aspects of Counts VI and VII, so The Times cannot disavow these prompts and outputs as they disavowed Exhibit J in order to avoid a waiver in connection with the copyright claim. If the parties are unable to resolve this issue, Microsoft will file a letter motion to compel in time for briefing prior to the October 30 conference.

The Times's Statement: Microsoft raised this dispute for the first time in a 7-page letter sent on the evening of October 15, and Microsoft has not yet responded to The Times's request to meet and confer on this dispute later this week. The Court's previous ruling on this point was not limited to the preparation of Exhibit J. Even so, the examples in the Complaint illustrate the types of conduct that underlie The Times's claims, and substantiate The Times's good faith basis for alleging that Microsoft is engaging in that conduct. The Times has no intention of relying on those specific examples at trial after it has had the opportunity to develop a fulsome record of Microsoft's conduct through discovery. Microsoft appears to be relitigating the same issues that OpenAI raised in its May 23, 2024 motion to compel privileged and work product documents and communications relating to the outputs cited in the Complaint and The Times's the pre-suit investigation. *See* Dkt. 124. The Court denied OpenAI's motion at the September 12, 2024 discovery hearing, at a hearing that of course Microsoft was present at and participated in. The Times's position on those issues is the same. If Microsoft claims that the issues they are raising are different, and the Court's prior ruling does not control, The Times remains willing to meet and confer with Microsoft to understand how. Until then, this issue is not ripe for the Court's adjudication.

2. The Times' ESI Custodians

OpenAI's Statement: Per the Court's guidance at the September 12 hearing, OpenAI has been diligently making its way through the Times's productions to determine whether to seek designation of additional custodians. To date, the Times has designated 12 custodians (as compared to OpenAI's 24 custodians), and there are clear holes in its designation of custodians with respect to its harm allegations and the Times's discovery of OpenAI's alleged infringement. For instance, the First Amended Complaint alleges "[m]isappropriation of [c]ommercial [r]eferrals" focused on Wirecutter content and recommendations, and claims that such alleged misappropriation "endangers Wirecutter's reputation" and "usurp[s] specific commercial opportunities." Dkt. 170 at 48-52, 66. Yet, the Times

has not identified a single Wirecutter custodian—OpenAI has proposed **Ben Frumin** (Wirecutter Editor-in-Chief) and **Jason Chen** (Wirecutter Editorial Director) to cure this issue. So too, the Times alleges that OpenAI’s products “threaten[] to divert readers, including current and potential subscribers, away from the Times,” yet has not proposed a custodian knowledgeable regarding subscription trends and analyses. OpenAI has proposed the addition of **David Gurian-Peck** (VP, Subscriber Monetization; Chief of Staff to CFO) to cure this issue. Finally, documents the Times has produced to date indicate that **Barrett Sheridan** (VP, Strategy and Operations), who was leading a team on behalf of A.G. Sulzberger regarding the Times’s generative AI strategy, is likely to have knowledge regarding the Times’s initial discovery of OpenAI’s alleged infringement. Absent the Times’s agreement to produce discovery from these custodians, OpenAI will seek relief from the Court.

The Times’s Statement: OpenAI provided the basis for seeking these four additional custodians for the first time via email on October 14 and asked The Times to respond by October 17. The Times is considering OpenAI’s request and plans to respond by the deadline provided. At the present time, The Times believes the parties can reach agreement for at least some of these custodians. The Times is happy to meet and confer with OpenAI to discuss any remaining areas of disagreement. Until then, this issue is not ripe for adjudication.

3. Date Ranges of The Times’ Searches for Responsive Documents

OpenAI’s Statement: The Times has confirmed that it is applying the date range of November 1, 2022 to May 1, 2024 to the majority of the terms it is running to search for responsive documents (the date range the Times has proposed for documents related to its AI Initiatives program is narrower: November 12, 2023 to May 1, 2024). This is plainly insufficient; the Times has made allegations of infringement going back to 2018, and has demanded discovery into earlier GPT models (GPT, GPT-1, GPT-2) on that basis. Cutting off its searches in late 2022 prejudices OpenAI’s ability to discover documents related to its statute of limitations defense. The Times has also made sweeping assertions regarding the alleged harm that it has suffered to its reputation, subscription numbers, and revenue, and yet seeks to deprive OpenAI of the discovery that would allow it to test those assertions. The Times should apply a date range of January 1, 2018 to July 1, 2024 to its search terms.³

Here too, the Times’s summary of the meet and confer history is inaccurate. To the extent the Times contends that OpenAI agreed to an overall date range of November 1, 2022 to May 1, 2024 for all the Times’s productions, that is incorrect. OpenAI and the Times agree that the parties should continue to meet and confer

³ To the extent a different range is required for specific terms, OpenAI will note this as necessary in briefing.

with respect to this issue; if the parties cannot resolve their dispute, OpenAI will seek relief from the Court.

The Times's Statement: OpenAI raised this issue for the first time on October 15, and has failed to confer in good faith on this issue as required by Local Rule 37.2 and this Court's Standing Order § II(b). The parties have been negotiating The Times's search terms since June 2024, and at every point so far, OpenAI has agreed to The Times's overall date range of November 1, 2022 to May 1, 2024.⁴ The Times circulated its most recent search term proposal (with these date ranges) on August 2, 2024. At no point until now has OpenAI asked The Times to expand the date range on any of these search terms. OpenAI even proposed the November 1, 2022 to May 1, 2024 date range for the 16 additional search terms it proposed. On October 12, 2024, OpenAI's counsel asked over email whether The Times would agree to search for responsive documents dating back to January 1, 2018 for certain specifically enumerated RFPs, many of which are being fulfilled through non-custodial searches, not search terms. On the parties' meet and confer call on October 14, The Times's counsel agreed to OpenAI's 2018 offer for 9 of those RFPs (including RFPs 35 and 37, which OpenAI discusses below). In response, OpenAI's counsel retracted the 2018 proposal, said the list of RFPs they sent on October 12 to which the proposal applied was incorrect, and said they would revert with a corrected list of RFPs for which they are seeking the 2018 date range. Instead of sending that list, OpenAI served this draft position statement on October 15 asking for the first time to expand the date range on every search term. The Times is happy to meet and confer with OpenAI regarding this new request, but to date OpenAI has not articulated a sufficient basis for this extremely broad request. Many of the search terms that OpenAI now wants The Times to apply on custodial documents dating back to January 2018 are directed at Times documents and communications regarding the harm OpenAI's ChatGPT tool has caused The Times—an issue that no one at The Times was discussing or documenting before ChatGPT was publicly released on November 30, 2022 (after the start of The Times's custodial document search term date range). *See* Dkt. 52 at 9.

4. The Times' Responses to OpenAI's RFPs and Interrogatories

OpenAI RFP Nos. 35 & 37

OpenAI's Statement: RFPs 35 and 37 request documents and communications related to (1) analyses relating to trends in readership or online subscriptions, and the causes thereof; and (2) analyses of the performance of content published on the Times's website, including analyses of methods to evaluate that

⁴ The Times applied a narrower date range of December 1, 2023 to May 1, 2024 for the custodial documents of Times employee Zach Seward and search terms relating to the "A.I. Initiatives Program" because Mr. Seward began working at The Times in December 2023 and launched the A.I. Initiatives Program shortly thereafter. Mr. Seward has no custodial documents relating to The Times's A.I. Initiatives Program predating December 2023 because the program did not yet exist and he was not yet employed by The Times.

performance. The Times initially refused to produce any documents in response to these requests, and ultimately agreed only to produce documents “sufficient to show”: (1) trends in readership for the Asserted Works and online subscriptions from 2018 through the present; (2) trends in revenue generated in connection with content published on the Times’s website from 2018 to present. Such productions are insufficient. The documents requested are clearly relevant to the Times’s allegations that OpenAI’s products have caused it harm. Evidence that neither readership nor online subscriptions declined during the period when OpenAI’s products have been publicly available, for example, would undermine the Times’s claims of harm. Similarly, evidence that content performance on the Times’s website was unaffected during that time period would undermine those claims. Documents “sufficient to show” such “trends” fail to provide the detail required for OpenAI to challenge the Times’s assertions. On the contrary, OpenAI must have access to the source material—the analyses themselves—in addition to whatever communications surrounding those analyses shed light on their validity. The very premise that documents “sufficient to show” only “trends” could be sufficient is flawed. For example, analyses of trends in readership could include analyses of various variables, such as demographic analyses (e.g., youth readership vs. adult readership) versus financial analyses (bundled subscriptions vs. news subscriptions). Different analyses could also reach different conclusions. Production of limited “trend” documents is insufficiently detailed to allow OpenAI to probe such details, and withholding communications surrounding such analyses improperly blinkers OpenAI’s defense.

The Times’s recitation of the meet and confer history of these requests is inaccurate. On August 6, OpenAI indicated that it was “open” to narrowing RFP 35 and “willing” to narrow RFP 37 to “documents sufficient to show,” so long as the Times would agree to produce such analyses for a lookback period of 30 years. The Times rejected that timeframe and no compromise was reached. OpenAI stands on its request that the Times conduct a search for documents and communications responsive to these requests, though it remains amenable to further meet and confer regarding the applicable timeframe. To the extent the parties cannot reach agreement, OpenAI will seek relief from the Court.

The Times’s Statement: The Times remains confused by OpenAI’s position on RFPs 35 and 37. The “sufficient to show” language for RFPs 35 and 37 was a compromise offer *made by OpenAI* during a July 31 meet and confer call and in an August 6 letter, which The Times then agreed to in writing on August 2 and again on August 14. OpenAI offered to narrow RFP 35 to “documents sufficient to show trends in readership for The Times’s works and online subscriptions covering The Times’s Asserted Works for some specified timeframe.” The Times agreed to produce documents responsive to this revised request from 2018 to present. Similarly, OpenAI offered to narrow RFP 37 to “documents sufficient to show trends in revenue generated in connection with content published on The Times’s website for some specified timeframe.” The Times agreed to that proposal as well, also for the date range of 2018 to present. Now, over two months later, OpenAI apparently

wishes to retract both offers entirely and takes issue with the language OpenAI itself proposed. OpenAI should be held to its own compromise offer, The Times's acceptance of which has now been in effect for two months, as this agreement appropriately limited these otherwise overly broad and unduly burdensome requests.

As for the date range of 2018 to present for RFPs 35 and 37, again it is OpenAI that offered this range, via an email to The Times on October 12. The Times accepted this offer on the parties' October 14 meet and confer call, but OpenAI then claimed its October 12 email was incorrect and it was seeking documents responsive to both requests going back to 2010. The Times understood that both sides were going to confirm whether they were amenable to meeting somewhere in the middle between 2010 and 2018. Since OpenAI does not mention any timeframe dispute in the above position statement, The Times assumes the parties are still negotiating the timeframe element. The Times remains amenable to additional meet and confers about the timeframe of The Times's responses to these two RFPs, although without additional discussion it maintains that OpenAI should be held to its October 12 proposal.

OpenAI RFP No. 70

OpenAI's Statement: RFP 70 requests, among other things, documents and communications concerning the benefits of GPT Services to the Times or journalism generally. The Times is limiting its production to the impact of GPT Services on the Times as alleged in the Complaint, but refusing to produce documents related to the impact on journalism more broadly. Such a narrow production prejudices OpenAI's defenses, particularly with regard to fair use. GPT Services have already had dramatic and positive impacts on many industries, providing capabilities in various marketplaces—including in journalism—that did not exist before. OpenAI is entitled to discovery related to the transformative effects of its technology in the broader journalism market, which are related at least to Factor 1 of OpenAI's fair use defense, in addition to broader market benefits afforded to news organizations by that technology, which are related to Factor 4. Limiting production to only the impact on the Times improperly deprives OpenAI of access to evidence that goes directly to its central defenses. Indeed, in meet and confer, the Times indicated that it still plans to argue that GPT Services have had a deleterious effect on the journalism market in general, while it simultaneously seeks to withhold evidence that would permit OpenAI to challenge that argument.⁵

The Times's Statement: Any relevant documents concerning "benefits" of OpenAI's copying of Times works are subsumed by what The Times already agreed to produce. *See, e.g.*, OAI RFPs 1, 3, 16, 17. To the extent that OpenAI seeks discovery into Times communications regarding other potential "benefits" of GPT

⁵ OpenAI further notes that, as with many other responses to its RFPs, the Times has refused to produce certain documents in reliance on the reporter's privilege. OpenAI reserves the right to continue to meet and confer on this issue and to present it to the Court as appropriate.

Services to “journalism generally” that do not relate to OpenAI’s infringing use of Times works or the effect of that use on the market, that material is irrelevant.

Broad discovery into all of The Times’s internal communications regarding any purported “benefits of GPT Services” to “journalism generally” is overbroad and outside the scope of fair use or any other defense in this case. Both the “transformative use” and “effect on the market” fair use inquiries are centered on *OpenAI’s infringing use* of Times works. The “effect on the market” analysis considers “the public benefits” of OpenAI’s use of Times works on the market and is driven by evidence like the “impact” of OpenAI’s copying on “on potential licensing revenues.” See *Hachette Book Group, Inc. v. Internet Archive*, 11 F. 4th 163, 192 (2d Cir. 2024). The “transformative use” factor focuses on the *character* of the infringing use. *Id.* at 179-180. The Times has already agreed to produce all documents and communications regarding OpenAI’s copying and unauthorized use of The Times’s works (see, e.g., OAI RFPs 1, 3, 16, 17). Moreover, the Times did not “indicate[] that it plans to argue that GPT Services have had a deleterious effect on the journalism market in general, while it simultaneously seeks to withhold evidence that would permit OpenAI to challenge that argument.”

OpenAI RFP Nos. 76 & 82

OpenAI’s Statement: RFP 76 requests documents and communications regarding any study, survey, or other research from any source regarding the market for the Times’s published works or the impact of Generative AI on the journalism market, from January 1, 2019 to the present. RFP 82 seeks documents and communications regarding applications of artificial intelligence large language models or applications for the purpose of improving the Times’s services, for marketing research, to personalize marketing, or to create audiences for third party advertisers.⁶ The Times has improperly refused to produce any documents in response to these RFPs, which are central to OpenAI’s fair use defense. As with the requests discussed above, OpenAI is entitled to evidence demonstrating the transformative nature of its technology, which is relevant at least to a defense of fair use, in addition to evidence demonstrating that OpenAI’s technology serves the public good without harming the Times. With regard to RFP 76, documents demonstrating that artificial intelligence large language models have benefitted the journalism market in any way would both support OpenAI’s defenses and undermine the Times’s claims of harm. The Times should not be permitted to simply refrain from searching for and producing materials that might undermine its claims. Similarly, with regard to RFP 82, the Times’s use of artificial intelligence large language models in its own operations—whatever the specifics of that use—would undermine the Times’s claimed harm. Such evidence would also illustrate

⁶ Terminology in RFP 82 differs from the terminology in RFP 76 because RFP 82 quotes the Times’s privacy policy, which refers to “artificial intelligence large language models”

how OpenAI's technology, far from damaging the Times's business, has provided new and transformative means by which the Times is seeking to grow.

The Times's Statement:

- **RFP 76:** As The Times said on the parties' October 14 meet and confer, The Times has agreed to produce documents sufficient to identify the market for providing paid access to The Times's copyrighted works (Microsoft RFP 57) and documents sufficient to identify the lost market value of the copyrighted works as a result of the conduct alleged in the Complaint (Microsoft RFP 58). That covers a large swath of what OpenAI seeks in RFP 76 regarding "the market" for Times works. As for the second half of RFP 76 seeking all documents concerning "any study, survey, or other research" regarding "the impact of Generative AI on the journalism market" generally, this Request is overbroad and unduly burdensome. The impact of third-party generative AI products on the overall journalism market is irrelevant to the fair use inquiry regarding *OpenAI and Microsoft's* infringing use of Times works in their own generative AI tools. *See* Dkt. 238 at 2; 17 U.S.C. § 107(1), (4). Nor are studies regarding third-party generative AI relevant to The Times's claims of harm by OpenAI and Microsoft. The Times already agreed to produce all documents and communications regarding OpenAI's use of Times works in OpenAI's generative AI tools (*see, e.g.*, OAI RFPs 1, 3, 16, 17) and documents regarding the harm OpenAI caused The Times (*see, e.g.*, OAI RFPs 26, 93; Microsoft RFP 56). To the extent there are any relevant documents regarding the "transformative nature" of OpenAI's infringing use of Times works, its effect on the market, or the harm it has caused The Times, those are already being produced.
- **RFP 82:** This dispute is covered by OpenAI's pending motion to compel filed on September 3, 2024. Dkt. 236. As set forth in The Times's September 6, 2024 opposition to that motion (Dkt. 238), The Times has agreed to substantial discovery into The Times's use of both Defendants' generative AI tools. *See id.* at 1. Discovery into The Times's use of third-party generative AI tools—especially those tools used "for marketing research, to personalize marketing, or to create audiences for third party advertisers"—is not relevant to whether OpenAI's infringement of The Times' works constituted fair use or caused harm. *See id.* at 1-2. The "transformative use" and "effect on the market" inquiries of fair use in this case relate to OpenAI and Microsoft's infringing use of Times works, not to the entire field of generative AI generally. *See* 17 U.S.C. § 107(1), (4). The Times's damages claims are likewise focused on the harm caused by OpenAI and Microsoft.

5. Status of Daily News Plaintiffs' Compliance With Their ESI and Document Discovery Obligations

In light of the Court's request at the last status conference for an update in this agenda on where things stand with respect to discovery, and given the relevance of this information to the schedule in this case, OpenAI provides the following information regarding the status of the Daily News Plaintiffs' collection, review, and production efforts.

a. The Daily News' Search Terms and Custodians

OpenAI's Statement: The ESI Order states that "the Producing Party will in the first instance propose search terms and custodians." See Dkt. 251 ¶ 5. The Daily News Plaintiffs have not yet proposed search terms and custodians.

Daily News Plaintiffs' Statement: Defendants (OpenAI) raised this issue for the first time yesterday. *Daily News* Plaintiffs responded to OpenAI's email this morning and notified them that *Daily News* Plaintiffs expect to produce search terms, custodians, and hit counts to Defendants by early next week. Nonetheless, Defendants insisted on raising this issue with the Court. As *Daily News* Plaintiffs told OpenAI in our email of this morning, we are currently in the process of running proposed search terms that correspond to Defendants' document requests across our entire custodial document collection. Before running such searches it was necessary to undertake the time-consuming task of gathering the electronic documents. We plan to provide Defendants with our proposed search terms and corresponding hit counts by early next week, and after the parties agree on the terms, we expect to begin a rolling production of custodial documents. This timing is consistent with how long it took for OpenAI (three and a half months) to provide search terms in its case against *The Times*.

b. The Daily News' Productions

OpenAI's Statement: The Daily News Plaintiffs have produced a total of 428 documents to date. The Daily News Plaintiffs have not produced any documents from any of the individual plaintiffs in this case.

Daily News Plaintiffs' Statement: Defendants' statement is misleading. The first 186 "documents" in *Daily News* Plaintiffs' production are spreadsheets comprising collated information on over 4 million asserted works, including the physical text of those over 4 million asserted works and links to the electronic versions of the over 4 million asserted works, as well as other information (e.g., dates, authors). *Daily News* Plaintiffs have also produced copies of their AI policies and have started to produce copyright registration certificates that correspond to the copyright registrations listed in the exhibits to Complaint. *Daily News* Plaintiffs will continue with a rolling production of additional copyright certificates and other non-custodial documents that are being collected and processed.

Defendants' statement that the *Daily News* Plaintiffs have not produced any documents of the individual plaintiffs in this case is also misleading. As *Daily News*

Plaintiffs explained to OpenAI by email this morning, the plaintiffs in the *Daily News* case are collectively managed by MediaNews Group (“MNG”). As such, non-custodial documents have been produced with the custodial designation MNG—which is where many of the documents reside. The documents themselves make clear to which entity they pertain (e.g., copyright certificates for Mercury News or Denver Post). When custodial documents are produced, *Daily News* Plaintiffs will identify from which custodian’s files those documents are being produced.

6. OpenAI’s Opposition to the News Plaintiffs’ Disclosure of Highly Confidential Material to Dr. Baeza-Yates

OpenAI’s Statement: The Daily News Plaintiffs have given notice of their intent to disclose to Dr. Ricardo Baeza-Yates, pursuant to ¶ 16 of the Protective Order, Dkt. 129, documents OpenAI has designated Highly Confidential under the Protective Order. OpenAI objects to the disclosure due to Dr. Baeza-Yates’s role as co-founder of and Chief Scientific Officer for Theodora.ai—a company that develops “AI-based tools.” OpenAI intends to file the requisite letter regarding its objection on Friday, October 18, 2024, as contemplated by ¶ 16(b) of the Protective Order. See Dkt. 129. Based on publicly available information and the information provided by Plaintiffs during the parties’ conferral, Theodora.ai attempts to innovate in the field of fine-tuning foundation models for specific tasks like bias-detection. OpenAI, in turn, has spent years developing industry-leading fine-tuning techniques like “reinforcement learning from human feedback (RLHF)” and reward modeling. Much of OpenAI’s work on that front is highly confidential and closely guarded. Disclosure of such highly confidential OpenAI material to Theodora.ai’s founder and current Chief Scientific Officer, who remains actively involved in the business, will allow Dr. Baeza-Yates to learn about OpenAI’s proprietary methods, which inherently creates “a risk of harm” that does not outweigh Plaintiffs’ need to disclose the Protected Material. See Dkt. 129 ¶ 16(b). Notably, Plaintiffs have not explained why they have a need to disclose OpenAI’s highly confidential information to Dr. Baeza-Yates beyond stating that he is a “retained consultant” with whom they seek to share OpenAI’s highly confidential documents and source code.

News Plaintiffs’ Statement: OpenAI’s opposition to disclosing its highly confidential information to Dr. Baeza-Yates because he is a cofounder and officer of a company that “develops AI tools” is unworkably broad. In a case where Large Language Models are at issue, it is simply unreasonable for OpenAI to oppose disclosing its highly confidential information to Plaintiffs’ technical expert because he works in the field. Under the Protective Order, OpenAI “bear[s] the burden of proving that the risk of harm that the disclosure would entail (under the safeguards proposed) outweighs the recipient’s need to disclose the Protected Discovery Material to its expert.” See Protective Order ¶ 16(b). The OpenAI Defendants have not met this burden.

During a recent meet and confer, the *Daily News* Plaintiffs asked OpenAI's counsel to provide any concrete, specific concerns it may have with disclosing highly confidential information to Dr. Baeza-Yates under the Protective Order. OpenAI's counsel was not able to articulate any such harm. As Plaintiffs explained to OpenAI's counsel, Theodora.AI—a company primarily focused on detecting bias in Spanish Language text—is not a competitor to OpenAI. If OpenAI submits a letter objecting to disclosure to Dr. Baeza-Yates, the News Plaintiffs will respond.

B. Issues Raised by Plaintiffs

The News Plaintiffs respectfully include the below list of disputes that have materialized and/or become ripe since the September 12 conference, which have not yet been before the Court. If the parties are unable to resolve these disputes in advance of the October 30 conference, the News Plaintiffs will file letter motions to compel.

1. Microsoft's ESI Custodians

News Plaintiffs' Statement: The parties have been discussing Microsoft's ESI custodians since early July, and have concluded their discussions with the exception of two disputes.

The *first* dispute concerns Microsoft employees Scott Counts and Sathish Manivannan, whom Microsoft identified in an Interrogatory response as knowledgeable about Microsoft's "efforts to track users' interactions with Defendants' Generative AI Products and Services, including use of Defendants' Generative AI Products and Services by The Times and the public." The Times initially requested that both Counts and Manivannan be added as document custodians, but now proposes the following compromise. Microsoft should select one of these employees to add as a custodian, while agreeing not to rely on the other's testimony at trial as well as in support of or in opposition to any summary judgment motion. This compromise alleviates the burden of collecting and producing additional documents, while ensuring the News Plaintiffs are protected against a situation in which Microsoft affirmatively relies on a witness whose documents were never produced.

The *second* dispute concerns Microsoft employee Mustafa Suleyman. Suleyman joined Microsoft in March 2024 as the CEO of its AI division. According to a recent report, he is "looking under the hood at OpenAI's crown jewels — its secret algorithms behind foundation models like GPT-4."⁷ Suleyman will therefore have unique documents about Microsoft's collaboration with OpenAI to train the models. Suleyman has also spoken publicly about his view while "anyone can copy" content "on the open web," there is a "separate category" that is a "gray area" where "a website or publisher or news organization had explicitly said, 'do not scrape or

⁷ <https://www.semafor.com/article/06/14/2024/microsoft-ai-ceo-mustafa-suleyman-audits-openais-code>.

crawl me for any other reason than indexing me.”⁸ Suleyman will therefore also have unique documents relevant to willfulness. Because Suleyman only recently joined Microsoft, any search of his files is unlikely to be burdensome. Moreover, the News Plaintiffs, as a compromise, have proposed that Microsoft apply a more limited set of search terms to his documents, and have made a specific proposal and requested hit counts.

Microsoft’s Statement: As with other issues, Microsoft and the News Plaintiffs have continued to meet and confer and work collaboratively regarding ESI custodians. To that end, Microsoft will continue to work with the News Plaintiffs its proposed compromise with respect to Scott Counts and Satish Manivannan as custodians, and does not expect there to be any dispute for the Court to resolve regarding this issue.

Microsoft however disagrees that Mustafa Suleyman is an appropriate custodian in this case. Mr. Suleyman was not employed by or otherwise affiliated with Microsoft when The Times filed its original complaint in December 2023, and did not join the company until March 2024, several months after the allegations in the complaint. As a result, The Times cannot credibly suggest that Mr. Suleyman actually has any custodial documents regarding any of the matters contained in the allegations in the complaint, given that he was not involved with Microsoft during those times. The allegations in The Times’ complaint involve the training of the GPT models, as well as allegations of harm resulting from outputs of Copilot and ChatGPT. Mr. Suleyman was not involved in either the training of the GPT models referenced in the complaint, or in the relevant aspects of Microsoft’s relationship with OpenAI. Likewise, he was not involved with the development or commercialization of Copilot as alleged in the Times’ complaint—again, because he was not affiliated with Microsoft during the development of Copilot. Instead, the Times is “headline hunting” based on what it asserts are Mr. Suleyman’s personal views (and at least the Semafor article references internet gossip not sourced to either Microsoft or OpenAI)—its only justification for nevertheless designating this individual as a custodian for purposes of ESI document collection rests on selected quotes from interviews given in June of 2024, which was nearly 6 months after the Times filed its complaint in this case. This is simply not a proper basis to designate Mr. Suleyman as a custodian for purposes of ESI document collection.

2. Microsoft Responses to The Times’s RFPs

News Plaintiffs’ Statement: The News Plaintiffs request that Microsoft be ordered to produce documents in response to RFP 67, which seeks documents concerning business development plans, strategic roadmaps, and product development plans related to (i) Defendants’ Generative AI Products and Services and (ii) any future products or services in the categories of news, cooking, consumer

⁸https://www.theregister.com/2024/06/28/microsoft_ceo_ai/#:~:text=%22I%20think%20that%20with%20respect,That's%20been%20the%20understanding.%22.

reviews, and sports. Such information is relevant to fair use and the substitutive effect of Defendants' products on journalism, among other issues.

Microsoft initially responded that it would not produce documents in response to this request. The parties met and conferred regarding this request on August 21 and 26, and Microsoft did not appear to move from its position. Nevertheless, News Plaintiffs attempted to confirm Microsoft's position in writing on August 29 and followed up regarding the same on September 17, September 25, and October 7, but never received a response. On October 10, News Plaintiffs asked whether Microsoft considered the newly announced Copilot Daily feature,⁹ which gives users a spoken summary of weather and current events, to be a new product that fell within RFP 67, or whether Microsoft would be producing documents regarding this product. The parties met and conferred regarding this issue on October 16 but did not reach a resolution.

News Plaintiffs will continue conferring with Microsoft in advance of the October 30 conference to reach agreement on this issue. If the parties are unable to resolve this dispute, the News Plaintiffs will file a letter motion to compel. These are not the only outstanding disputes related to Microsoft's RFP responses, but the News Plaintiffs respectfully seek to prioritize this issue for the Court's consideration at the upcoming October 30 conference.

Microsoft's Statement: Microsoft and the Times have met and conferred extensively regarding a wide range of Microsoft's responses to the Times' interrogatories and requests for production of documents. As a result of those meet and confers, the parties were able to resolve a number of issues, and Microsoft has agreed to supplement many of its responses, which touch on nearly all issues in these cases. Since that time, Microsoft has also continued to investigate the issues raised by the Times, and as a result, Microsoft communicated to the Times that the process of supplementation has taken longer than initially anticipated. Microsoft expects to serve these supplemental responses before the October 30 conference.

However, with respect to Request No. 67, the parties have not reached a resolution. This request on its face does not seek information regarding the products and services actually alleged in the Times' complaint, but instead seeks wide-ranging information regarding Microsoft's *future* products and services. This includes products that have not been commercially released, products that may be in development, and even plans for products that may ultimately not be released or even developed. In fact, by seeking documents regarding "any future products or services in the categories of news, cooking, consumer reviews, and sports," the Times is also seeking Microsoft's future product development information that is not even related to Generative AI in any way. As such, this Request is wholly outside the scope of reasonable discovery and is an effort to seek wide-ranging information regarding Microsoft's development pipeline. Should the Times continue

⁹ <https://blogs.microsoft.com/blog/2024/10/01/an-ai-companion-for-everyone/>

to pursue discovery in this case regarding Microsoft’s future products and plans, Microsoft anticipates that this issue will require more formal briefing.

3. The Scope of OpenAI’s Products Subject to Discovery

News Plaintiffs’ Statement: News Plaintiffs used the defined term “Generative AI Product(s) and Service(s)” across a number of discovery requests, including requests for production, requests for inspection, and interrogatories. News Plaintiffs defined this term to mean “any publicly available or commercial product or service offering that includes a Generative AI Model, including but not limited to ChatGPT branded products and services, Copilot branded products and services, ChatGPT, Browse with Bing, Bing Chat, OpenAI’s soon-to-be announced search product, APIs, and platforms otherwise providing access to a Generative AI Model.” In its response, OpenAI objected, and limited the scope of the documents that it intends to produce only to ChatGPT.

These products and services are relevant to both The Times’s case and the Daily News case because OpenAI used the News Plaintiffs’ content to develop these products. Indeed, the *Daily News* Complaint expressly identifies several generative AI products and services in addition to ChatGPT, including Copilot, Browse, CustomGPTs, OpenAI’s APIs, and all other platforms otherwise providing access to ChatGPT and/or the GPT models. See *Daily News* Compl. ¶¶ 13–16, 59, 63–71, 114–38, 147–52. Given these allegations, and the direct, vicarious, and contributory copyright infringement counts included in the Complaint, News Plaintiffs submit that it is improper for OpenAI to limit the scope of products to which News Plaintiffs’ discovery requests pertain just to ChatGPT. OpenAI has previously argued that providing discovery into all of these products would be overly burdensome, but those concerns are outweighed by the relevance of this discovery. OpenAI should not be permitted to use the scope of its misconduct to evade relevant discovery.

During a recent meet and confer (October 14), OpenAI’s counsel tentatively proposed including ChatGPT and, additionally, Copilot, Browse, CustomGPTs, and OpenAI’s APIs, with additional products to be considered as they are identified, and their relevance shown. OpenAI’s counsel indicated that they would take this offer back to OpenAI to determine if it was something with which OpenAI could agree, and that their main concern was having a fixed set of products to work with. The News Plaintiffs will continue to meet and confer with OpenAI in an effort to resolve this issue.

OpenAI’s Statement: As the News Plaintiffs acknowledge, the parties continue to meet and confer regarding the scope of products subject to discovery. Consequently, there is no dispute that currently requires the Court’s attention.

OpenAI has never refused to expand the definition of the term “Generative AI Product(s) and Services(s).” To the contrary, it has repeatedly expressed that it

would be willing to consider broadening the definition—even to products that Plaintiffs have never identified in any of their complaints. The issue still being negotiated involves OpenAI’s request that the list of products and services be discrete and reasonable in number.

OpenAI has attempted to ascertain the specific products Plaintiffs understand to be included within the definition of “Generative AI Product(s) and Service(s).” Contrary to Plaintiffs’ recounting of the parties’ latest meet and confer, OpenAI’s counsel did *not* propose including an open-ended list of products within the term’s definition. Rather, counsel asked Plaintiffs to confirm a proposal of a concrete, limited set of products OpenAI could consider.

Absent such limits, discovery would be unworkable. Take, for example, Plaintiffs’ demand that OpenAI include “CustomGPTs” in the definition of “Generative AI Product(s) and Services(s).” As their name implies, CustomGPTs are tailored versions of ChatGPT *created by third party users* that are customized for very specific purposes, from helping users improve their creative writing, providing troubleshooting advice for electronics, or quickly explaining board game rules. These CustomGPTs can be created by any ChatGPT user; in other words, they are not created by OpenAI. And there are a plethora of them—at present, roughly **3 million**.

Needless to say, responding to discovery requests encompassing millions of products customized by third party users would be onerous. The Times’ Interrogatory No. 10, for example, would require OpenAI to identify, in writing, all versions and release dates of every one of the 3 million CustomGPTs that are currently available. And responding to RFPs would be just as burdensome. For example, The Times’ RFP No. 35 asks for “[d]ocuments concerning any changes made to Defendant’s . . . Generative AI Products and Services from 2021 to the present.” In other words, to respond to the request as Plaintiffs demand, OpenAI would have to produce millions of documents addressing minor, likely inconsequential changes, to millions of different CustomGPTs.

There is no conceivable theory of relevance that would justify that burden. OpenAI has already agreed to produce discovery related to the defined universe of OpenAI **models** that power ChatGPT (and each of these millions of third party user products). And it is OpenAI’s **models** that Plaintiffs allege were trained on Plaintiffs’ content. *See, e.g., Daily News Compl.* ¶ 3. Plaintiffs have not identified any basis that all of these third party user products are relevant to their claims.

OpenAI remains willing to work with Plaintiffs to identify the OpenAI products and services that are most relevant to its claims and will continue to try to do so. But that effort must reflect reasonable limits to make Plaintiffs’ requests proportional to the needs of the case.

4. OpenAI's Responses to RFPs and Interrogatories

News Plaintiffs' General Statement: The News Plaintiffs request that OpenAI be ordered to produce documents in response to the below set of RFPs. These are not the only outstanding disputes related to OpenAI's RFP responses, but the News Plaintiffs respectfully seek to prioritize these RFP responses for the Court's consideration at the upcoming October 30 conference.

For many of these RFPs, it would appear from the below write-ups that the parties have not met issue. The reason is this. The News Plaintiffs sent OpenAI its position statements, as articulated herein, on October 15, pursuant to an agreed upon exchange deadline. OpenAI has since responded, including with some willingness to move off its previous positions, though unfortunately in most cases not enough. The News Plaintiffs will continue to meet and confer with OpenAI in advance of the October 30, hearing, in an effort to further narrow the disputes. If the parties are able to further resolve any issues, they will so update the Court in advance of the October 30 hearing. Alternatively, to the extent the parties are unable to resolve them, the News Plaintiffs will file a letter motion to compel.

OpenAI's General Statement: OpenAI has been working cooperatively with the News Plaintiffs to resolve disputes related to the following RFPs and interrogatory. It will continue to do so in advance of the October 30 conference. OpenAI notes that these RFPs are belated additions to this agenda. Each of the RFPs raised by the News Plaintiffs was the subject of correspondence sent by OpenAI to the New York Times on August 20. The New York Times did not respond until more than 6 weeks later—in emails on October 4 and 9—just days before this agenda was due. As a result, the parties have been unable to meet and confer about their current positions on these requests. Nevertheless, OpenAI provided updated responses to these requests on October 15 and 16, and continues to discuss the substance in good faith.

New York Times RFP 18:

News Plaintiffs' Statement: This RFP seeks documents relevant to The Times's DMCA claims—specifically, documents “concerning the removal of any content, including information regarding authorship, date of publication, publishing entity, title, copyright notices, and terms and conditions for use of works, from Training Datasets before, during, or after training of Defendants' Generative AI Model(s).”

Until October 16, OpenAI had not yet committed to producing any documents in response to this RFP, notwithstanding its core relevance to The Times's DMCA claims. See FAC ¶¶ 184-86 (alleging that Defendants violated sections 1201(b)(1) and (b)(3) of the DMCA by “remov[ing] The Times's copyright management information (“CMI”) in building the training datasets containing millions of copies

of Times works,” and by “remov[ing] The Times’s [CMI]” in the process of “generating outputs” from Defendants’ user-facing products like ChatGPT).

OpenAI’s Statement: This request, seeking “documents concerning the removal of any content . . . from Training Datasets” (emphasis added), is overly broad, unduly burdensome, and disproportionate to the needs of the case. As written, this RFP could, for example, sweep in ordinary data cleansing steps like deleting formatting or typos. In recent meet and confer communications, the New York Times explained for the first time that it is seeking documents relevant to its DMCA claim. In light of that clarification—and contrary to the Times’ assertion—OpenAI has agreed to produce documents in response to this request. Specifically, OpenAI has agreed to conduct a reasonable search for and produce non-privileged documents in its possession, custody, or control discussing whether to remove, from datasets on which the relevant models were trained, the following categories of information with respect to copyrighted works: authorship, date of publication, publishing entity, title, copyright notices, and terms and conditions for use.

New York Times RFPs 26-27:

News Plaintiffs’ Statement: These requests seek documents concerning OpenAI’s knowledge of and use of web crawlers and related tools to access Times content, such as but not limited to GPTBot, ChatGPT-User, OAI-SearchBot, CommonCrawl bot, as well as any crawlers used by Microsoft or developers of Custom GPTs. RFP 26 seeks documents “concerning Your use or knowledge of web crawlers, bots, spiders, user agents, and related tools to access The Times’s Content, including Times websites and digital products.” And RFP 27 seeks documents “concerning Your reliance on, use of, or understanding of third-party robots.txt crawler directives or third-party meta tag crawler directives.” These requests are relevant to, at a minimum, The Times’s claim of willful infringement. FAC ¶¶ 124-26.

Until October 16, OpenAI’s sole offer for RFP 26 was to make training data available for inspection. That response is insufficient. While training data may shed light on which content was used to train the models, it may not reveal *how* that content was accessed and obtained. And for RFP 27, OpenAI has merely offered to produce documents “sufficient to show OpenAI’s approach to robot.txt, as explained by [public] OpenAI blog posts.” This response is also insufficient. OpenAI should search for and produce non-public documents in response to this request, including custodial documents.

OpenAI’s Statement: These requests seek documents related to web crawlers and related tools to access New York Times content. In recent meet and confer communications, the Times clarified that these requests are intended to seek “documents concerning OpenAI’s knowledge of and use of web crawlers and related tools to access Times content.” But the Times has also said that these requests seek

“documents concerning the operation of such crawlers” without limitation. To the extent these requests seek such documents, they are vague and overbroad.

Nonetheless, in the interest of compromise, for RFP 26, OpenAI has agreed to conduct a reasonable search for and produce non-privileged documents in its possession, custody or control concerning OpenAI's use of web crawlers to crawl web content of articles published by the Times in connection with training the relevant models. With regard to RFP 27, OpenAI has agreed to produce responsive, non-privileged documents in its possession, custody, and control, if any, sufficient to show OpenAI's approach to robots.txt.

The Times has not offered any explanation as to why OpenAI's offers are insufficient. OpenAI remains willing to meet and confer on these requests to better understand the Times's position.

New York Times RFPs 44, 53:

News Plaintiffs' Statement: These RFPs seek documents concerning how Defendants' products circumvent paywalls. *See* RFP 44 (seeking documents “concerning any attempts to use Defendants' Generative AI Products and Services to circumvent paywalls or other access control measures”); RFP 53 (“Documents concerning the decision to suspend Browse with Bing and whether or not to suspend Bing Chat”). RFP 53 specifically targets the OpenAI's publicly reported decision in July 2023 to suspend its Browse with Bing feature over concerns about how the product enabled users to evade paywalls.¹⁰

For RFP 44, OpenAI until October 16 only agreed to produce documents “sufficient to show ChatGPT's approach to paywalls”, which is insufficient because it does not address the substance of the RFP, which targets users' conduct and, moreover, it excludes custodial documents. For RFP 53, OpenAI until October 16 refused to produce any documents. Both of these requests are relevant to The Times's harm allegations; The Times alleges that Defendants' products enable users to circumvent The Times's paywall to access copyrighted Times content, and that Defendants' products in this way harm The Times. FAC ¶¶ 156-57. These allegations also support The Times's contributory infringement claim, which seeks to hold Defendants accountable to the extent an “an end-user may be liable as a direct infringer based on output of GPT-based products.” FAC ¶ 179. OpenAI argued in its motion to dismiss that this claim should fail because “normal people do not use OpenAI's products” to elicit copyrighted content. Dkt. 52 at 2. It is inappropriate for OpenAI to make fact-bound arguments about how people use Defendants' products, only to refuse to produce documents relevant to those assertions.

OpenAI's Statement: The Times describes these requests as seeking documents related to “attempts to use Defendants' Generative AI Products and

¹⁰ https://www.theregister.com/2023/07/05/openai_pauses_bing_search/.

Services to circumvent paywalls.” Contrary to the Times’s statement, OpenAI has agreed to produce documents in response to both requests.

In response to RFP 44, OpenAI agreed to conduct a reasonable search for non-privileged documents in its possession, custody, or control discussing OpenAI's approach to paywalls with respect to the training data for the relevant models and products.

As to RFP 53, which seeks documents specific to a product called “Browse with Bing,” OpenAI has already agreed to conduct a reasonable search across its existing custodians and produce non-privileged documents in its possession, custody, or control, related to Browse with Bing’s interactions with paywalls. OpenAI has also agreed to produce documents, to the extent they exist, regarding how paywalls may have related to the decision to suspend Browse with Bing and/or Bing Chat, if at all.

OpenAI believes these offers resolve any dispute as to these requests, but has not yet heard back from the Times.

New York Times RFPs 45, 47-48, 51-52:

News Plaintiffs’ Statement: These requests seek documents relating to user engagement metrics for Defendants’ products, including documents concerning how OpenAI tracks and logs information about user queries and product outputs. RFP 45 seeks “query, session, and chat logs reflecting or analyzing user sessions related to Times Content within Defendants’ Generative AI Products and Services.” RFP 47 seeks documents “concerning user engagement metrics, events, and analytics data collected during each user session.” RFP 48 seeks documents “sufficient to show (i) the number of daily active users of each of Defendants’ Generative AI Products and Services, separated into the numbers of new users and repeat users on each day, from 2021 to present, and (ii) dates of subscription and changes in subscriber numbers over time.” RFP 51 seeks documents “concerning the behavior of users of Defendants’ Generative AI Products and Services, including [] documents concerning the extent to which users of Defendants’ Generative AI Products and Services navigate away from Defendants’ Generative AI Products and Services to other websites.” And RFP 52 seeks documents “concerning the extent to which Defendants’ Generative AI Products and Services direct users to Times Content in response to user queries.”

Until October 16, OpenAI had not committed to producing any documents in response to these requests, which are relevant to The Times’s damages. These requests are also relevant to fair use, insofar as they would reveal the extent to which Defendants’ user-facing products (e.g., ChatGPT) provide copyrighted news content as output to users.

OpenAI's Statement: These requests seek documents related to OpenAI's users' engagement, their queries, and related outputs. OpenAI is diligently investigating what responsive, nonprivileged information exists regarding each of these RFPs. Contrary to the Times' assertion, OpenAI has agreed to produce documents in response to several of these RFPs. OpenAI has also committed to supplement its responses as appropriate as its investigation continues. The Times has not yet responded to these proposals.

- **RFP No. 45:** RFP 45 seeks "Query, session, and chat logs reflecting or analyzing user sessions related to Times Content within Defendants' Generative AI Products and Services." OpenAI's investigation into these documents remains ongoing, and it has agreed to update the Times in due course.
- **RFP No. 47:** RFP 47 seeks documents "concerning user engagement metrics, events, and analytics data collected during each user session." In response, OpenAI has agreed to conduct a reasonable search for non-privileged documents in its possession, custody, or control analyzing or summarizing the rate at which OpenAI users click links included in responses provided by the relevant products.
- **RFP No. 48:** RFP 48 seeks documents "sufficient to show (i) the number of daily active users of each of Defendants' Generative AI Products and Services, separated into the numbers of new users and repeat users on each day, from 2021 to present, and (ii) dates of subscription and changes in subscriber numbers over time." OpenAI has agreed to conduct a reasonable search for non-privileged documents sufficient to show the number of daily active users of each of the relevant products, from 2021 to the present, in the manner and at the frequency such information is kept in the ordinary course of business.
- **RFP No. 51:** RFP 51 seeks documents "concerning the behavior of users of Defendants' Generative AI Products and Services, including [] documents concerning the extent to which users of Defendants' Generative AI Products and Services navigate away from Defendants' Generative AI Products and Services to other websites." OpenAI has agreed to conduct a reasonable search for non-privileged documents in its possession, custody, or control analyzing or summarizing the rate at which OpenAI users click links included in responses provided by the relevant products. OpenAI further agreed to conduct a reasonable search for non-privileged documents in its possession, custody, or control discussing OpenAI's efforts, if any, to modify clickthrough rates for the relevant products.
- **RFP No. 52:** RFP 52 seeks documents "concerning the extent to which Defendants' Generative AI Products and Services direct users to Times

Content in response to user queries.” OpenAI’s investigation into these documents remains ongoing, and it has agreed to update the Times in due course.

OpenAI believes its agreements have resolved the relevant disputes.

New York Times RFPs 74-81:

News Plaintiffs’ Statement: These requests seek documents concerning the “New York Times” datasets that OpenAI compiled for training their models. Documents produced in discovery reveal that OpenAI targeted Times content because of its “high quality,” and then “scraped” Times content for compilation into specific datasets. *E.g.*, Dkt. 204-2 (internal OpenAI document attached to prior motion to compel); Dkt. 228-4 (similar); Dkt. 228-6 (similar). *See* RFP 74 (“All documents relating to the NYT Datasets”); RFP 75 (“All documents relating to Your purchase, acquisition, or obtainment of the NYT Datasets”); RFP 76 (“All documents relating to any contract, license agreement, or terms applicable to the NYT Datasets”); RFP 77 (“All documents relating to Your ingestion or processing of the NYT Datasets”); RFP 78 (“All computer code or scripts used to ingest or process the NYT Datasets”); RFP 79 (“All documents relating to the use or contemplated use of the NYT Datasets”); RFP 80 (“All documents relating to any testing or evaluation You performed on the NYT Datasets”); RFP 81 (“All documents concerning any and all scrapes or downloads You performed of Times Content, including the ‘scrape’ of Times Content referenced in OPCO_NYT_0116390 and the seven million Times URLs referenced in OPCO_NYT_0432207”).

OpenAI has so far not agreed to produce any documents in response to these requests, instead contending that these requests are duplicative to prior requests. While prior requests do seek documents concerning the training process and data used to train the models, these requests are unique because they target specific conduct revealed through discovery. OpenAI should conduct a reasonable investigation to locate any additional documents in response to these requests, particularly given their obvious relevance.

OpenAI’s Statement: All of these requests seek documents related to OpenAI’s training datasets. As OpenAI has repeatedly explained to the Times, it appears the documents sought by these RFPs, with the possible exception of RFP No. 78, are duplicative of numerous prior RFPs which also seek documents related to OpenAI’s training datasets. *See, e.g.*, RFPs Nos. 1, 8, 12, 20, 21, 22, 23, 24, 28, and 49. But to the extent these requests break new ground, OpenAI has agreed to conduct a reasonable search for non-privileged documents in its possession, custody, and control describing whether and how the relevant models were trained using articles published by The Times, to the extent those documents exist. OpenAI believes that this offer resolves any potential dispute as to RFP Nos. 74-77 and 79-81.

For RFP No. 78, which appears to seek source code related to OpenAI's ingestion of training data, OpenAI has asked to meet and confer to better understand the information the Times is seeking with those requests.

New York Times RFPs 84-85:

News Plaintiffs' Statement: These requests seek documents concerning OpenAI's use of websites like DNYuz that copy publisher content without permission.¹¹ DNYuz is a website "filled with content plagiarized from publishers, including the Times, the Atlantic, Bloomberg, and the Financial Times, among others." RFP 84 seeks "all documents concerning Your knowledge of DNYuz (www.dnyuz.com), including its copying of Times Content and Your use of DNYuz content for any purpose." RFP 85 relatedly seeks "all documents concerning Your knowledge of websites besides DNYuz that regularly copy publisher content without permission, and Your use of such websites for any purpose."

OpenAI has so far refused to produce documents in response to these requests, which are relevant to, among other issues, The Times's claim of willful infringement. FAC ¶¶ 124-26. Documents in response to these requests will reveal whether OpenAI used DNYuz or similar websites to obtain publisher content without permission. Indeed, OpenAI has conceded the relevance of third-party websites that copy publisher content by stating publicly that regurgitation is "more common when particular content appears more than once in training data, like if pieces of it appear on lots of different public websites."¹²

OpenAI's Statement: Contrary to Plaintiffs' assertion, OpenAI has not "refused to produce documents in response to these Requests." OpenAI has already run "dnyuz" as a search term and reviewed the resulting hits. OpenAI has also already agreed to produce documents discussing (i) how and why OpenAI curates and uses different types of data to train OpenAI's relevant models; and (ii) the relative value of types of data, if any, for use in the relevant models. OpenAI has offered to meet and confer to understand what more The Times is seeking.

As for RFP No. 85, The Times has failed to identify any websites that are purportedly "similar" to DNYuz such that OpenAI could conduct a reasonable search for those websites. Nor has The Times explained how this request is proportional to the needs of the case.

¹¹ <https://www.buzzfeednews.com/article/craigsilverman/drudge-report-links-site-plagiarizes-stories>; see also <https://futurism.com/chatgpt-plagiarized-nyt-articles> ("In Futurism's testing, ChatGPT regularly cited DNYuz' plagiarism factory as an authoritative and original source. And even after we alerted OpenAI to the issue, ChatGPT continued basing its answers on DNYuz's stolen content.").

¹² <https://openai.com/index/openai-and-journalism/>

OpenAI remains eager to meet and confer to better understand The Times's position on these requests.

Daily News Interrogatory 2:

News Plaintiffs' Statement: The *Daily News* Plaintiffs' Interrogatory No. 2 asks Defendants to identify each instance of Publisher Content used, referenced, or accessed as part of Retrieval Augmented Generation methods for OpenAI's Generative AI Products and Services. This information is plainly relevant to the case, and Defendants are in the best position to identify this content, which is not readily accessible through other means of discovery.

During a recent meet and confer (October 14), OpenAI's counsel explained that OpenAI is still investigating what it can provide with respect to this interrogatory from a technical perspective. They proposed limiting the scope of their answer to instances where one of the *Daily News* Plaintiffs' URLs is used, referenced, or accessed as part of the Retrieval Augmented Generation methods, and will consider whether they can answer given this limited scope. The *Daily News* Plaintiffs will continue to meet and confer with OpenAI in an effort to resolve this issue.

During a recent meet and confer (October 15), Microsoft's counsel explained that Microsoft is still investigating what it can provide with respect to this interrogatory from a technical perspective. The parties discussed technical issues around the data, and possible ways to proceed, including whether limiting search of the data to a set of specific domains. The *Daily News* Plaintiffs will continue to meet and confer with Microsoft in an effort to resolve this issue.

OpenAI's Statement: OpenAI and the *Daily News* had a meet and confer about this interrogatory for the first time on October 14, 2024. OpenAI explained that the framing of the interrogatory may misunderstand how the technology works but it is willing to continue its conversation with the *Daily News* to provide an appropriate response. The parties' conferral was productive and considerably narrowed the issues. OpenAI is continuing to investigate its response and will continue to meet and confer.

Respectfully submitted,

/s/ Annette Hurst
Annette Hurst